

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/150825>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Becoming literate while learning a second language – practicing reading aloud

Catia Cucchiarini¹, Mario Ganzeboom¹, Joost van Doremalen², Helmer Strik^{1,2}

¹Centre for Language and Speech Technology, Radboud University, Nijmegen, the Netherlands

²NovoLanguage, Nijmegen, the Netherlands

c.cucchiarini | m.ganzeboom | w.strik @let.ru.nl, joost@novolanguage.com

Abstract

The DigLin project aims at providing concrete solutions for low-literate and illiterate adults who have to learn a second language (L2). Besides learning the L2, they thus also have to acquire literacy in the L2. To allow intensive practice and feedback in reading aloud, appropriate speech technology is developed for the four targeted languages: Dutch, English, German and Finnish. Since relatively limited resources are available for this application for the four studied languages, this had to be taken into account while developing the speech technology. Exercises with suitable content were developed for the four languages, and are tested in four countries: Netherlands, United Kingdom, Germany, and Finland. Preliminary results are presented in the paper, and suggestions for future directions are discussed.

Index Terms: adult literacy learning, language and speech technology, second language acquisition

1. Introduction

Skills like reading and writing are often taken for granted, esp. in western countries. However, there are many low-literate and illiterate people, even in western countries, who have to struggle to achieve these skills. According to UNESCO [27], about 775 million adults are illiterate, among which 122 million are young people. Many immigrant and refugee adults who arrive in Europe have a low education level and limited literacy. These people will have to learn to read and write in a language other than their mother tongue and will face the double task of becoming literate while at the same time acquiring a second language. It is well known that these learners encounter enormous difficulties in learning new languages [1] [2] [3] [4]. A compounding problem is that, in general, limited resources are available to support these learners in this difficult task. Financial resources are limited because many countries have cut down on adult education. As a consequence, learning materials for this specific target group are also limited. Language learning materials that are now becoming available on the internet, sometimes even for free, are not easy to find for learners that are not able to read and write. Another additional problem are cultural and social differences that sometimes constitute real barriers to education. Illiterate learners often feel ashamed and are reluctant to attend literacy courses.

Researchers and teachers have been looking for innovative solutions that can make literacy acquisition more effective, efficient, autonomous and motivating. The project ‘Digital Literacy Instructor’ (DigLin) funded by the Lifelong Learning Program (LLP) is such an initiative [5] [6]. DigLin aims at developing and testing innovative materials for adult literacy students. Some of the exercises employ Automatic Speech Recognition (ASR) to analyze the learner’s read speech output and provide feedback. This form of active practice in which

literacy students can produce the sounds or words while a computer tells them whether they are correct is a much needed improvement. There have been various initiatives in which ASR was employed in literacy acquisition [7] [8] [9] [10], but – as far as we know – this technique has not yet been applied in literacy education to adult second language learners.

The three year DigLin project started in January 2013. The partners in DigLin are:

- CLST, Radboud University Nijmegen (the Netherlands), coordinator [11];
- Friesland College (the Netherlands) [12];
- University Newcastle upon Tyne (United Kingdom) [13];
- University of Vienna (Austria) [14];
- University of Jyväskylä (Finland) [15].

2. The pedagogical approach in DigLin

In this project we depart from a common framework, digital sources of FC-Sprint² [16] [17], and develop content and exercises in keeping with the specific features and requirements of the language and the teachers in question [18]. The underlying method in FC-Sprint² [16] [17] and the one used in DigLin is in fact a phonics-based method: the structure method. The primary aim of the structure method is grasping the structure of the spelling system or associating specific sounds (phonemes) with specific letters (graphemes). This is done on the basis of a whole word which is visually and auditorily structured in smaller units (analysis). In this way the student learns to consider a written word as a composite unit of separate elements and to make use of the systematic nature of letter-sound associations for autonomously decoding new words.

The basis of this method is a restricted number of concrete basic words the meaning of which is clear. In classes of 6- and 7-year-old children, those words are presented in a context of a story or a picture story and learnt by heart. In DigLin those words can be made clear by pressing a button. Basic words should have a ‘one-on-one grapheme-phoneme correspondence’, that is to say that the pronunciation of the sounds is only influenced in a limited way by preceding or following sounds or by the fact that they are in word-final or syllable-final position, as is the case in Dutch. We use the label “pure sound”. Some examples for:

- English: dad, map, mop, jump, bin, big, yes
- Dutch: mat, kap, kip, boom
- German: Rat, Hut, Oma
- Finnish: eno, iso, akka

Ideally, there is a one-to-one relationship between phoneme- and grapheme. This is not always the case, since many languages have too few graphemes for the repertoire of phonemes, which is the case for Dutch, but more particularly for English with one and the same grapheme representing different phonemes.

As soon as a couple of basic words are recognized, the analysis and synthesis exercises can start. The spoken word is

analyzed in sounds, the written word in letters. Next, the sounds are blended to a spoken word. Many analysis and blending exercises are needed for establishing a tight association between sounds and letters. Software can help to automatize this phase of the reading process. For this stage, FC-Sprint2 has found many challenging exercises with feedback (e.g., a letter dragged to an incorrect position, does not stay, but jumps away, back to its original position).

3. Automatic Speech Recognition in DigLin

ASR of non-native speakers can be challenging [19] especially in the case of illiterates [20] and in the case of beginner L2 learners [21]. In the DigLin project speech technology has to be developed for low- or illiterate people that are beginner learners of a second language, which thus constitutes a very challenging task.

While for many languages databases of native speech are available, corresponding databases of non-native speech are in general lacking. The languages involved in DigLin are Dutch, English, German and Finnish. For these target groups (low- or illiterate beginner L2 learners) very limited resources are available. This makes it even more challenging to develop speech technology for this application. In DigLin, we cope with this issue in the following way. We start with an ASR trained on native material, using native resources (lexica, speech corpora, etc.). We then study whether using extra information can improve the system's performance, e.g. by using non-native resources (lexica, speech corpora, etc.), and by using information on errors made by the target group (annotations of errors). The limited available non-native audio recordings and error annotations are first used, while interactions of users with (initial versions of) the system, and annotations of (part of) these recordings will be employed at a later stage to improve the system.

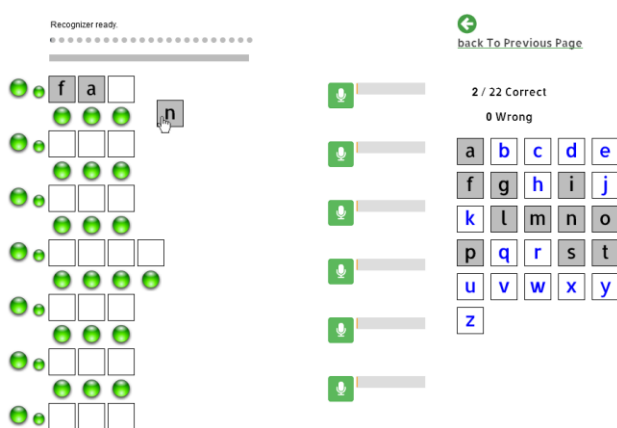


Figure 1. Screenshot of the 'Drag the letters' exercise in FC-Sprint².

Exercises have been developed such that the possible answers by the users are restricted. Figure 1 shows an example of such an exercise. In this exercise learners are presented with an example pronunciation of a word by clicking on the left most green, marble button. They then have to identify and drag the letters of the words into the slots behind the button. Learners can get a visual hint to the word when they hover over the smaller green, marble button and hear the pronunciation of the individual letters when clicking on the marble buttons below the slots.

For every item, a list of correct and incorrect responses is used to limit the recognition task. The DigLin system is intended to be web-based, and should run in different browsers. Since practical, technical details can be important for a good performance, we carefully looked at issues such as head-sets, audio recording settings (for different browsers), audio file formats, signal-to-noise ratio (SNR), and noise cancelling (techniques).

In general, feedback should be intuitive, and easy to interpret. This is especially the case for the current target groups. We have been experimenting with different possibilities, discussed them with experts, and in the end decided to use the following set-up. When the pronunciation of a word is not correct, feedback is provided to signal this to the learner. Feedback is gradual in the sense that it indicates the degree of correctness. A student can repeat again and again and a slider indicates in real time whether there is any improvement so that the student can try again immediately and see whether the new attempt is better or worse.

Learners can also listen to correct examples in stored audio recordings. Students can repeatedly listen to these example speech recordings in the program, as often as they want. When making these audio recordings we carefully considered criteria such as speed, accuracy of pronunciation, amount of silence, whether or not carrier sentences should be used, good selection of speakers (male and female, amount of dialect, etc.), recording environment and conditions (studio, 'silent office'), technical specifications (e.g. file format (wav/mp3), signal-to-noise ratio (SNR), etc.). Eventually, we decided to present the speech in the program at normal speed instead of slow speed so as to prevent a stark contrast between the slow speech usually employed by teachers to real world speech.

4. Method

Speech recognition

In this project, we use the Speech Recognition and Automatic Annotation Kit (SPRAAK) [22], an open source semi-continuous Hidden Markov Model (HMM) ASR package. The input speech, sampled at 16 kHz, is divided into overlapping 32ms Hamming windows with a 10 ms shift and pre-emphasis factor of 0.95. 12 Mel Frequency Cepstrum Coefficients (MFCCs) plus C0, and their first and second order derivatives were calculated and Cepstral Mean Subtraction (CMS) was applied. The constrained language models and pronunciation lexicons are implemented as Finite State Transducers (FSTs). Three-state, context-independent acoustic models with a left-to-right topology were trained for all languages involved. For Dutch and English the well-developed Spoken Dutch Corpus (Corpus Gesproken Nederlands, CGN) [23] and Wall Street Journal (WSJ) [24] corpora were already available. These also provide segmentations (for part) of the training material to bootstrap the training of the acoustic models. For German and Finnish we used the SpeechDat-Car corpora [25]. Initial segmentations for the German and Finnish SpeechDat-Car corpora were obtained by using Dutch acoustic models. In order to do so, we created mappings between the Dutch phone set and the German and Finnish phone sets. The resulting segmentations were used to obtain bootstrap acoustic models for these two languages.

Finite State Grammars (FSG) were used as language models. This FSG allowed one or multiple instances of the target word in order to model repetitions of words, and optional filled pauses/silences in order to model hesitations.

Recording learner speech

To test the performance of the speech recognition system, recordings of learner speech were required for all languages. Project partners recorded learner speech by having students read out the prompting words from the set of language exercises. At the start of the project we made an inventory of which first languages (L1s) are most relevant for the four countries involved in DigLin. For the four target languages (L2s) involved, the resulting recordings of the partners contain audio files for the L1s that were indicated to be most important. Table 1 provides details on the number of speakers and recordings per target language.

Target language	Num. of speakers	Num. of recordings
Dutch	25	6839
German	17	4530
English	18	6533
Finnish	17	4832

Table 1. Number of speakers and recordings per target language.

The transcribed recordings were used in a word recognition task to test the performance of the speech recognition systems for the different languages. In this task, normalized acoustic likelihoods were calculated as confidence scores. For each of the recordings, the confidence score was determined for the target word and another randomly chosen word from the set of words used in the exercises. Distributions of the confidence scores that a word was correctly or incorrectly recognised were derived for every language. Subsequently, the equal error rates (point at which the number of false positives and false negatives are equal, EER) were calculated to investigate the discriminative ability of the confidence score.

Providing feedback to the learner

In the web-based system, feedback on the learner's speech was implemented through a visual slider (see Figure 2). This is done in the following way. First, we determine if 0, 1, or N occurrences of words are spoken. If no target words are recognized feedback is provided that the target word is not recognized, and if N words are recognized the feedback is that multiple words are recognised. In both cases (0 or N words recognized) the slider shows a score of 0. Only in the case that 1 target word is recognized a score between 0 and 1 is calculated. Figure 2 shows a screenshot of the visual slider implementation in DigLin.

To determine a value between 0 and 1 the confidence score of the recognised word was scaled. The scale used is based on a sigmoidal function as described in the results section.

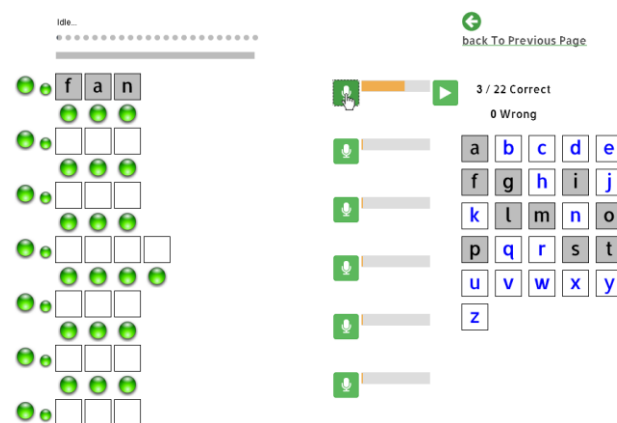


Figure 2. Screenshot of the feedback given to the learner after a pronunciation attempt of the word 'fan'.

5. Results

The word recognition task described in the previous section resulted in two sets of confidence scores: scores of the speech aligned with the target word and scores of the alignment with another randomly chosen word from the set of words used in the exercises. For instance, Figures 3, 4 and 5 show kernel density estimates of the histograms of these sets for Dutch, German, and Finnish, respectively. The horizontal axis shows the normalized acoustic likelihoods (i.e. confidence scores) and the vertical axis the number of words.

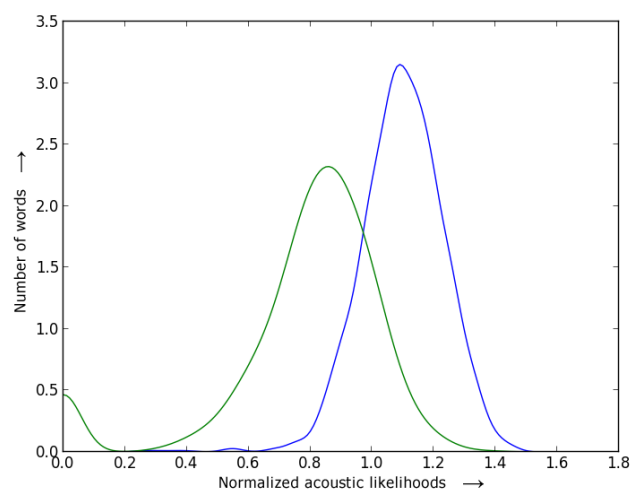


Figure 3. Kernel density estimates of the confidence scores for Dutch. Blue shows the scores of the audio aligned with the target word and green the score when aligned with another randomly chosen word.

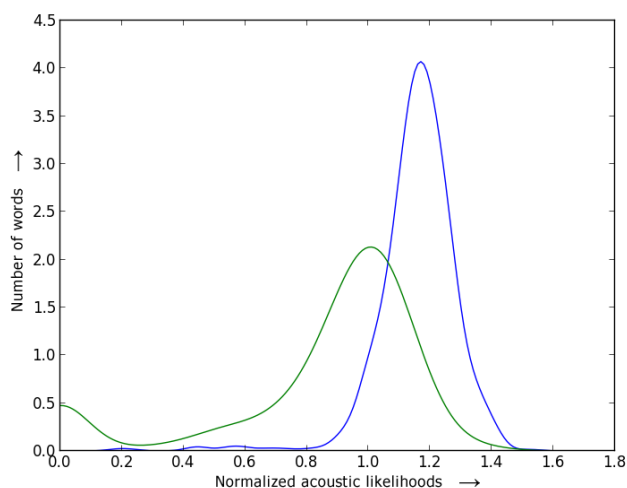


Figure 4. Kernel density estimates of the confidence scores for German. Blue shows the scores of the audio aligned with the target word and green the score when aligned with another randomly chosen word.

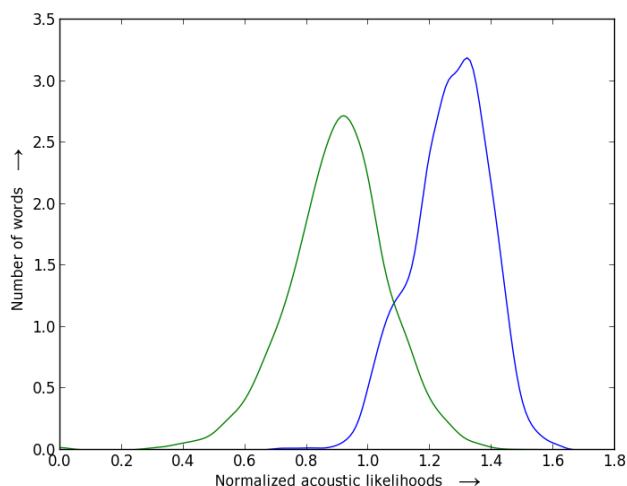


Figure 5. Kernel density estimates of the confidence scores for Finnish. Blue shows the scores of the audio aligned with the target word and green the score when aligned with another randomly chosen word.

In the ideal case, the two distributions of confidence scores do not intersect. Which is equal to the system being 100% confident about its true positives and negatives. In the worst case, the distributions intersect fully. Where the blue and green line intersect the confidence of the system is 50% for either case. As the figures show, all blue lines are for the larger part to the right of the green ones. This shows that in general the system does provide a higher probability for having recognised the target word in comparison to that of the random other word. However, there is also some overlap.

The next issue, was to calculate a suitable score that could be used to provide feedback to the learners. This was done in the following way. Suppose that the likelihood ratio between a correct and incorrect pronunciation is $N:1$, then the feedback score is $N/(N+1)$. For example, when the chance of a correct versus incorrect pronunciation is 1:1 (i.e. point where the blue and green lines intersect in Figure 3 - Figure 5), the output score is $1/(1+1) = 0.5$. In the case that the ratio is 4:1, the score is $4/(4+1) = 0.8$. Such a relation is modelled by a sigmoid function and is shown in red for Finnish in Figure 6. In Figure 6 it can be observed that at the point where the green and blue lines cross, where the ratio is 1:1, the resulting score is 0.5, and that at the right of this crossing point the score becomes larger, increasing to 1, and at the left of this crossing point the score becomes smaller, decreasing to 0.

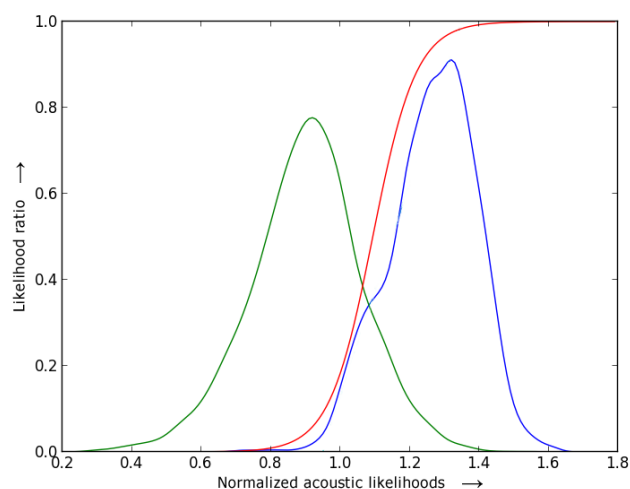


Figure 6. Values of Finnish with the corresponding likelihood ratio on the y axis, modelled by a sigmoid function (red line).

Figures 3 to 5 also show that for the Dutch and German distributions the amount of overlap seems to be similar, while for Finnish the amount of overlap is smaller. This is also reflected in the EERs, which are 17%, 18.5%, 10.9% for Dutch, German, and Finnish, respectively. The difference between the EERs of Dutch and German compared to that of Finnish is notable. A possible explanation might be the higher transparency of Finnish orthography (this is one of the reasons why Finnish was chosen as one of the languages in the DigLin project) and the corresponding more direct grapheme-phoneme correspondences, which could make the task in Finnish less complex providing better results with the same amount of data.

28NED - 2305
8129 show_word_picture's
8137 hide_word_picture's
4550 play_word_sound's
3829 play_letter_sound's
503 play_soundbar_word_sound's
469 session_stop's
7606 letter_drag's
6731 letter_drag_right's
877 letter_drag_false's
272 exercise_end's
428 asr-start's
225 asr-result's
1271 word_drag's
5002 word_drag_false's
1224 word_drag_right's
2339 input_focus's
2331 input_blur's
3 play_soundbar_word_sound_gray's

Exercises	1	min	2	min	3	min	4	min	5	min	6	min	7	min	8	min	9	min	10	min	11	min	12	min	13	min	14	min	15	min
The words	3	20	8	0	0	0	0	0	0	0	0	0	0	0	3	6	3	8	2	1	5	10	1	2	3	2	5	18	4	13
Drag the letters	12	32	4	8	4	10	5	11	8	29	6	20	7	58	10	22	22	82	15	77	21	107	27	91	27	121	32	169	15	84
Listen and drag words	0	0	1	3	1	3	2	6	3	10	2	10	0	0	0	0	3	4	3	10	0	0	1	3	2	7	5	19	7	29
Form and drag the words	5	16	4	9	1	2	2	5	3	8	3	7	6	14	2	6	4	11	5	10	9	20	8	24	10	27	7	23	9	39
Listen and type	1	0	0	0	2	1	0	0	2	15	0	0	0	0	0	0	0	0	1	6	1	6	1	2	1	0	1	6	2	1
Read the words	14	54	4	30	5	43	6	46	4	31	5	46	4	22	11	55	9	98	21	134	9	82	8	107	4	43	4	51	16	98
Test yourself	0	0	0	0	1	1	2	0	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Figure 6. An example of a visualization of some of the information present in the log-files. Shown is an overview of the activity of one learner (with the code 28NED), which has used the DigLin system for 2305 minutes. In DigLin there are 7 types of exercises (incl. “Test yourself”), and for each type of exercise there are 15 versions with different content of increasing complexity. The table presents an overview of how often each exercise was done, and how many minutes were spent on it. Above the table is some other information regarding the behavior of this learner.

Besides the experiments mentioned above, we conducted some ad hoc experiments to test the quality of the speech technology developed with the procedures described above. In general, the outcomes of these experiments were positive. However, for German we noticed some segmentation problems, especially for word initial fricatives. A possible reason could be that the German acoustic models are trained with the SpeechDat-Car corpus, simply because this corpus was available at the start of DigLin. SpeechDat-Car corpora were collected for research on speech recognition in a car environment. Considering that this is a noisy environment, the recordings also contain a certain level of noise. This differs from the (generally) less noisy office environments in which DigLin is used, and that could be a reason segmentation problems were observed for German. At the moment, we are investigating this by training acoustic models on the German SpeechDat corpus [25]. The recordings in this corpus better match the ‘silent office’ environment and thus acoustic models trained using the SpeechDat corpus might yield better segmentations.

6. Discussion and Conclusions

The DigLin system has been developed for the four languages involved, and is currently being evaluated in four countries: the Netherlands, United Kingdom, Germany, and Finland. Results of these evaluations will be presented at the SLaTE workshop. All interactions of the users with the DigLin system are stored in log-files, and the spoken utterances are stored on the ASR server. These data (log-files and audio files) provide

a rich source of information. Tools have already been developed to visualize certain aspects of the log-files. An example is presented in Figure 6. These tools are, e.g., used by the researchers in the different countries to keep track of the activities of the learners, since it can be seen which exercises were carried out, in which order, how often, how much time it took the learners, etc. Obviously, log-files and audio files also can be used for other research purposes. Audio files as training and testing material to improve the speech technology modules. Log-files to get a better idea about the learning behaviour, to observe what the successful and less successful components of the DigLin system are, and thus how the DigLin system can be improved. Results of these analyses will be presented.

Preliminary results are encouraging. In general, the DigLin system seems to function well, and teachers’ impressions are that many learners have already made substantial progress. ASR also seems to constitute a valuable add-on for many exercises. For the first time, this makes it possible for learners to receive automatic, immediate feedback on their spoken utterances. These low-literate and illiterate adults, e.g., have to learn to make letter to sound correspondences, how words can be broken up in individual sounds (analysis), and how individual sounds can be combined to form words (synthesis). This learning process can be improved, if they can speak, and get feedback on it.

Preliminary analyses also revealed some issues that might need further attention. An important issue is that these learners can read words in many different ways. In our language model, we already took into account that multiple words could

be spoken (instead of 1 target word), and that there could be silences or filled pauses. However, in reality the situation is much more complex. For instance, there are also other disfluencies, broken words, and these learners often read 'letter by letter', probably because they have problems reading the whole word. The question then is what to do with all these different ways of reading. An option is to keep the language model as it is, and then the learners should simply speak correctly, i.e. read 1 target word with a (fairly) correct pronunciation, and they should keep trying to do so until the feedback tells them that their utterance was correct. Another option is to try to improve the language model, to better model the different ways of reading. However, it is not immediately clear what the benefits might be. With an improved language model it might be possible to provide more detailed feedback, but teachers and other experts doubt whether this is useful for these learners. All these issues provide interesting thoughts for further research.

In any case, what has become clear is that ASR can be valuable for low-literate and illiterate adults learning a second language. The nature of the exercises, the language tasks involved is such that constrained ASR tasks can be designed, which in turn makes it possible to obtain adequate ASR performance. And by using ASR they can practice speaking in the L2, while receiving immediate feedback. This is an important improvement for L2 reading instruction, which paves the way to more autonomous learning conditions.

7. Acknowledgements

This project has been funded with support from the European Commission under project: 527536-LLP-1-2012-1-NL-GRUNDTVIG-GMP. This publication reflects the views only of the project consortium members, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

We are indebted to the other members of the DigLin team for their contributions, in alphabetical order: Ineke van de Craats, Marta Dawidowicz, Jan Deutekom, Enas Filimban, Vanja de Lint, Maisa Martin, Rola Naeb, Jan-Willem Overal, Karen Schramm, Taina Tammelin-Laine, and Martha Young-Scholten.

8. References

- [1] Onderdelinden, L., I. van de Craats & J. Kurvers (2009). Word concept of illiterates and low-literates: worlds apart? In I. van de Craats & J. Kurvers (eds.) *Low-Educated Adult Second Language and Literacy Acquisition*, 4th Symposium - Antwerp 2008: Utrecht: LOT Occasional Series 15, 35-48.
- [2] Strube, S. (2014). Grappling with the oral skills. The learning and teaching of the low-literate adult second language learner. Utrecht: LOT.
- [3] Tammelin-Laine, T. (2011). Non-literate immigrants – a new group of adults in Finland. In C. Schöneberger, I. van de Craats & J. Kurvers (eds) *Low-Educated Adult Second Language and Literacy Acquisition*, 8th Symposium – Cologne 2010: Nijmegen: Centre for Language Studies, 67-78.
- [4] Young-Scholten, M., & Naeb, R. (2009). Non-literate L2 adults' small steps in mastering the constellation of skills required for reading. In T. Wall and M. Leong (Eds.). *Low-Educated Second Language and Literacy Acquisition: Proceedings of the 5th Symposium*, Banff, 2009, 80-81.
- [5] <http://hstrik.ruhosting.nl/diglin/>
- [6] <http://www.diglin.eu/>
- [7] Duchateau J. Kong, Y. Cleuren, L. Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière, P., Verhelst, W., Van hamme, H. (2009). Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules. *Speech Communication* 51(10): 985-994.
- [8] Mostow, J., Roth, S., Hauptmann, A.G., Kane, M., A (1994). Prototype Reading Coach that Listens. *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI)* 785-792.
- [9] Russell, M., D'Arcy, S. (2007). Challenges for computer recognition of children's speech. *Proc. SLATE-2007*, pp. 108-111.
- [10] Li, Y., and Mostow, J. (2012). Evaluating and improving real-time tracking of children's oral reading. In *Proceedings of the 25th Florida Artificial Intelligence Research Society Conference (FLAIRS-25)*, Marco Island, Florida.
- [11] <http://www.ru.nl/clst/>
- [12] <http://www.frieslandcollege.nl/>
- [13] <http://www.ncl.ac.uk/>
- [14] <https://www.univie.ac.at/en/>
- [15] <https://www.jyu.fi/en/>
- [16] Deutekom, J. FC-Sprint², Grenzeloos Lernen, Boom 2008.
- [17] <http://www.fcsprint2.nl/>
- [18] Cucchiari, C., Craats, I. van de; Deutekom, J.; Strik, H. (2013) The digital instructor for literacy learning. *Proc. of the SLATE-2013 workshop*, Grenoble, France, pp. 96-101.
- [19] Benzeghiba, M., R. D. Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouviet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens (2007) Automatic speech recognition and speech variability: a review, *Speech Communication*, vol. 49, no. 10-11, pp. 763-786.
- [20] Al-Barhamtoshy, H., Abdou, S. and Rashwan, M. (2014) Mobile Technology for Illiterate Education, *Life Science Journal*, 11(9), 242-248.
- [21] Doremalen, J.J.H.C. van, Cucchiari, C. & Strik, H. (2010). Optimizing automatic speech recognition for low-proficient non-native speakers. *EURASIP Journal on Audio, Speech and Music Processing*.
- [22] Kris Demuynck, Jan Roelens, Dirk Van Compernelle and Patrick Wambacq. SPRAAK: An Open Source Speech Recognition and Automatic Annotation Kit. In *Proc. International Conference on Spoken Language Processing*, page 495-498, Brisbane, Australia, September 2008.
- [23] Oostdijk, N. (2002) The design of the Spoken Dutch Corpus, in Peters P., Collins P., Smith A. (Eds) *New Frontiers of Corpus Research*, Rodopi, Amsterdam, 105-112.
- [24] Douglas B. Paul and Janet M. Baker. 1992. The design for the wall street journal-based CSR corpus. In *Proceedings of the workshop on Speech and Natural Language (HLT '91)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 357-362. <http://dx.doi.org/10.3115/1075527.1075614>
- [25] Moreno, A., Borge, L., Christoph, D., Gael, R., Khalid, C., Stephan, E., Jeffrey, A., 2000. SpeechDat-Car: a large speech database for automotive environments. In: *Proc. II LREC*.
- [26] H. Hoge, H. Troph, R. Winski, H. van den Heuvel, R. Haeb-Umbach, "European Speech Databases for Telephone Applications", ICASSP, 1997, Acoustics, Speech, and Signal Processing, IEEE International Conference on, Acoustics, Speech, and Signal Processing, IEEE International Conference on 1997, pp. 1771, doi:10.1109/ICASSP.1997.598873
- [27] <http://www.unesco.org/new/en/education/themes/education-building-blocks/literacy/resources/statistics>